

ABSTRACT

A system retrieves and processes information stored on computers connected by a communications network. A central computer receives notification from a remote computer that the computer is available to receive data. In response to that notification, the central computer sends address data to the remote computer. To utilise the available network resources with maximum efficiency, the central computer optimises performance of the distributed system by allocating address data to the remote computer based on at least one characteristic of the remote computer, such as a measure of network connectivity and/or on a performance characteristic of that remote computer. This allocation may take place in accordance with the relative importance of the data for indexing purposes. The remote computer uses a communication interface connected to the Internet to retrieve the information stored at the locations specified by the address data, and stores that information. The remote computer then processes the retrieved information to generate processed data, and stores the processed data. Finally, the remote computer sends the processed data to the central computer.